

## BAB II

### TINJAUAN PUSTAKA

#### 2.1. PENELITIAN TERDAHULU

Pada penelitian yang dilakukan oleh beberapa peneliti dengan judul yang berbedadapat dijadikan bahan pertimbangan pada penelitian ini dan diharapkan mampu membantu.

Menurut Ronny Setiawan dengan penelitiannya yang berjudul “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru”. Penelitian ini membahas tentang penerapan data mining, menggunakan algoritma K-Means Clustering untuk menghasilkan profil yang memiliki kemiripan atribut yang sama. Metode yang digunakan adalah CRISP- DM dengan melalui proses business understanding, data understanding, data preparation, modeling, evaluation dan deployment. Implementasi proses K-Means Clustering menggunakan Rapid Miner 7.0. Atribut yang digunakan usia, agama, status ujian, kelulusan, registrasi, kelamin, gelombang pendaftaran, gelombang registrasi, Nilai Tes, Jurusan, program studi, sumber informasi, asal kota, lokasi kampus, pekerjaan orang tua, jenis sekolah, pembayaran uang registrasi. Menghasilkan jumlah kluster 4 ( $k=4$ ) dengan cluster pertama 17007 calon mahasiswa, cluster kedua sebanyak 83 calon mahasiswa, cluster ketiga sebanyak 12919 calon mahasiswa dan kluster keempat sebanyak 356 calon mahasiswa (Setiawan, 2017).

Menurut narwanti dalam penelitiannya dengan judul “pengelompokan mahasiswa menggunakan algoritma K-Means” yang membahas mengenai pengelompokan mahasiswa berdasarkan data akademik menggunakan teknik clustering algoritma K-means. maka dilakukan pengelompokan berdasarkan data nilai tes mahasiswa ketika masuk dan indeks prestasi kumulatif mahasiswa. Maka dengan pengelompokan data tersebut diperoleh hasil berupa informasi pola

prestasi mahasiswa yang clusternya tetap, naik atau turun bisa dilihat dari program studi dan sekolah asal mahasiswa (Narwati, 2012).

Menurut totok suprawoto yang penelitiannya berjudul klasifikasi data mahasiswa menggunakan metode K-Means untuk menunjang pemilihan strategi pemasaran. dalam penelitian ini peneliti menggunakan clustering data mahasiswa dengan metode k-means, data yang diambil terdiri dari nilai rata-rata ujian nasional dan indeks prestasi kumulatif mahasiswa. Dengan pengujian tersebut diperoleh hasil pengujian yaitu kelompok mahasiswa dengan nilai rata-rata ujian nasional yang rendah memiliki pengaruh terhadap prestasi akademik mahasiswa yang rendah pada jenjang diploma -3(D3) dan strata -1(S1-). Jika seorang mahasiswa memiliki nilai ujian nasional yang tinggi maka prestasi akademik mahasiswa juga tinggi pada semua jenjang. Sehingga dengan hasil pengelompokan berdasarkan daerah asal sekolah IPK rata-rata yang tertinggi berasal dari provinsi daerah istimewa Yogyakarta dan Jawa Tengah(Suprawoto, 2016)

Dalam penelitian rima dias ramadhani yang berjudul “data mining menggunakan algoritma k-means clustering data mahasiswa baru dengan algoritma k-means”. Pada penelitian ini atribut yang digunakan adalah asal sekolah, jurusan dan IPK siswa yang kemudian di cluster menjadi 3 cluster. Kemudian dengan hasil penelitian ini digunakan sebagai salah satu dasar pengambilan keputusan untuk kemudian di jadikan strategi promosi berdasarkan cluster yang terbentuk oleh pihak akademik di UDINUS (Ramadhani, 2013)

Dalam Penelitian Eka Sabna yang berjudul “Analisis Data Mahasiswa Dengan Algoritma K-Mean Untuk Mendukung Strategi Promosi Stikes Hang Tuah Pekanbaru”. Pada penelitian ini bertujuan untuk mengklasifikasikan data Mahasiswa Baru STIKes Hang Tuah Pekanbaru dengan menggunakan Teknik Data Mining. Teknik Penambangan Data yang digunakan adalah teknik Clustering dengan algoritma K-Means dan untuk proses menggunakan metode CRISP-DM. Hasil pengelompokan data mahasiswa melalui persebaran wilayah berdasarkan Program Studi

dan Kota Asal menggunakan K-Means clustering terbentuk ga Cluster yaitu Cluster 2 dengan jumlah 367 mahasiswa yang berasal dari kota dengan kode 1-10, Cluster 1 dengan jumlah mahasiswa 221 yang berasal dari kota dengan kode 4-18 dan cluster 0 dengan jumlah mahasiswa 63 orang yang berasal dari kota dengan kode 14-37 (MPUTER, t.t.).

## 2.2. LANDASAN TEORI

### 2.2.1 Promosi Dan Srategi Pemasaran

#### 2.2.1.1. Promosi

Menurut Cahsanah dan widiyono dalam kotler, *promotion, the fourth marketing mix tool, stands for various, the company undertakes to communicate it's product merits and to persuade target customers to buy them.* Pengertian ini mempunyai pengertian bahwapromosi meliputi semua alat yang terdapat dalam bauran promosi yang peranan utamanya adalah mengadakan komunikasi yang bersifat membujuk (Chasanah & Widiyono, 2017)

Menurut kotler dan Armstrong variable-variabel yang ada di dalam promotion mix ada lima yaitu sebagai berikut (Sukotjo & Radix, 2010):

#### a. Periklanan (*advertising*)

Periklanan adalah suatu bentuk atau cara membujuk orang dengan mengajak berkomunikasi guna mengetahui maksud dan tujuan kita. Pada umumnya promosi itu mencakup produk atau layanan yang di mana kita menawarkan agar bisa memberikan manfaat bagi konsumen. Tujuan khusus dari kegiatan promosi adalah membujuk, mempengaruhi, dan memberikan pemahaman kepada pelanggan tentang jasa dan produk yang kita promosikan.

b. Penjualan personal (*personal selling*)

Pada konsep penjualan yang dimaksud adalah bahwa seorang individu yang melakukan penjualan atau mempromosikan suatu produk dengan cara tatap muka dengan metode presentasi, Tanya jawab, guna mendapatkan pemesanan.

c. Promosi penjualan (*sales promotion*)

Promosi dalam penjualan yang dimaksud adalah kegiatan tatap muka dengan cara berkomunikasi dimana perusahaan bersama konsumen dapat mempengaruhi dalam pembelian sesuai kemauan dan kebutuhan.

d. Hubungan masyarakat (*public relation*)

Dalam kegiatan promosi kita dapat bersentuhan langsung dengan masyarakat banyak guna memperoleh kepercayaan pada sebuah produk sehingga efeknya positif dan bisa menciptakan hubungan kerja sama.

e. Pemasaran langsung (*direct marketing*)

Pemasaran langsung adalah sistem pemasaran interaktif yang menggunakan satu atau lebih media untuk mempengaruhi suatu tanggapan atau transaksi terukur pada lokasi mana pun”(Adila & Aziz, 2019)

Pengertian lain dari pemasaran langsung menurut Belch dan Belch yang dikutip oleh Kennedy (2006:26): “Direct marketing is a system of marketing by which organization communicates directly with the target consumer to generate a response or transaction”.

Sedangkan pada Saladin (2004:191) mengemukakan bahwa “Pemasaran langsung adalah suatu sistem pemasaran yang menggunakan suatu media iklan atau lebih guna mendapatkan respons dan/ atau transaksi yang bisa diukur di suatu lokasi”.(Adila & Aziz, 2019)

Promosi sebagai suatu kegiatan yang dilakukan oleh perusahaan untuk mengkomunikasikan manfaat dari produknya dan untuk meyakinkan konsumen agar membeli. (Philip kotler, 1997:142)

Promosi sebagai salah satu variabel dalam bauran pemasaran yang sangat penting dilaksanakan oleh perusahaan dalam memasarkan produk jasa(Selang, 2013).

Promosi sebagai serangkaian teknik yang digunakan untuk mencapai sasaran penjualan atau pemasaran dengan penggunaan biaya yang efektif, dengan memberikan nilai tambah pada produk atau jasa baik kepada para perantara maupun pemakai langsung, biasanya tidak dibatasi dalam jangka waktu tertentu. (Julian Cummins, 1991:11)

Dari pengertian atau definisi di atas bisa disimpulkan bahwa promosi merupakan suatu kegiatan yang tidak hanya dilakukan perusahaan dengan pembeli mengenai suatu produk maupun jasa tapi promosi juga digunakan sebagai wadah untuk menarik maupun mempengaruhi pembeli/konsumen pada kegiatan jual beli sesuai kemauan sebagai kebutuhan pokok.

#### **2.2.1.2. Strategi Pemasaran**

Strategi promosi merupakan suatu bentuk kegiatan yang intinya mengarah pada pencapaian dan tujuan dari strategi pemasaran yakni untuk mendapatkan pelanggan (*mind share*.)pada dunia strategi pemasaran di temukan tiga macam dimensi yang dimana untuk dapat membantu pencapaian tujuan. Adapun macam-macam dimensi strategi pemasaran yaitu (Rizal & Sandiana, 2016):

##### **1. Segmentasi**

Segmentasi adalah suatu metode yang dimana pasar akan di bagi berdasarkan variable-variabel tertentu sesuai dengan kondisinya

## 2. Targeting

Targeting merupakan bentuk metode yang akan melahirkan segmen-segmen dan kemudian akan di pilih satu atau dua segmen untuk dijadikan *target market*.

## 3. Positioning

Positioning pada dimensi strategi pemasaran dikenal dengan posisi, di mana penjelasan secara rincinya yaitu sebuah posisi yang dikejar oleh perusahaan untuk kemudian bisa mengetahui apa yang diinginkan konsumen.

### 2.2.2 Data Mining

Datamining adalah “proses yang menggunakan teknik statistika, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar”(Ramadhani, 2013)

data mining memiliki pengertian lain yaitu *knowledge discovery* ataupun *pattern recognition* merupakan suatu istilah yang digunakan untuk mendapatkan pengetahuan yang tersembunyi dari kumpulan data yang berukuran sangat besakvjr. Tujuan utama data mining adalah untuk menemukan menggali, atau menambah pengetahuan dari data atau informasi yang kita miliki (Rony setiawan, 2016).

Menurut Connolly dan Begg data mining adalah suatu proses ekstraksi atau penggalian data yang belum diketahuisebelumnya, namun dapat dipahami dan berguna dari database yang besar serta digunakan untuk membuat suatu keputusan bisnis yang sangat penting. sedangkan menurut vercellis, data mining adalah aktivitas yang menggambarkan sebuah proses analisis yang terjadi secara iteratif pada database yang besar, dengan tujuan mengekstrak informasi data knowledge

yang akurat dan berpotensi berguna untuk knowledge workers yang berhubungan dengan pengambilan keputusan dan pemecahan masalah (DIANTORO, 2016)

Dari banyaknya pengertian yang di ungkapkan maka peneliti menyimpulkan bahwa data mining adalah sebuah bidang yang sangat berkembang dengan cepat karena dengan besarnya nilai tambahnya dari database untuk pengukuran data yang ukurannya dalam skala besar sehingga sangat dibutuhkan. Dalam perjalanannya dengan ini juga didapatkan sebuah pola di mana sangat menarik yang tujuannya adalah mengekstrak sebuah informasi serta knowledge yang tepat, berpotensi, dan sangat mudah dipahami dalam mengambil keputusan yang sangat penting sehingga bisa dipergunakan.

### **2.2.3 Pengelompokan Data Mining**

Menurut Larose pengelompokan data mining terbagi atas berbagai kelompok namun berdasarkan tugasnya kemudian dapat dilakukan sebagai berikut (Menarianti, 2015):

a. Deskripsi

deskripsi yaitu “untuk mencari cara yang menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pada pola dan kecenderungan yang sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan”.

b. Estimasi

Estimasi bisa dikatakan sama dengan klasifikasi, yang membedakan adalah “variabel target estimasi lebih ke arah numerik daripada ke arah kategori”. Model di bangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

c. Prediksi

Prediksi bisa didefinisikan hampir sama dengan klasifikasi dan estimasi, kecuali bahwa “pada prediksi dalam memprediksi nilai akan di ambil dari hasil dimasa mendatang”.

d. Klasifikasi

Klasifikasi dalam klasifikasi yaitu “terdapat target variabel kategori Sebagai contoh penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah”.

e. Pengklusteran

Pengklusteran yaitu “pengelompokan record, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan”. Kluster adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidakmiripan dengan record dalam kluster lain. Algoritma pengklusteran mencoba untuk melakukan pembagian terhadap keseluruhan data menjadi kelompok-kelompok yang memiliki kemiripan, yang mana kemiripan record dalam satu kelompok akan bernilai maksimal, sedangkan kemiripan dengan record dalam kelompok lain akan bernilai minimal.

f. Asosiasi

Asosiasi dalam data mining memiliki tugas yaitu “menemukan atribut yang muncul dalam suatu waktu, dalam dunia bisnis lebih umum disebut analisis keranjang belanja”.

#### 2.2.4 Metode Data Mining

CRISP-DM adalah standarisasi data mining yang disusun oleh tiga penggagas data mining market. Yaitu Daimler Chrysler (Daimler-Benz), SPSS (ISL), NCR. Kemudian dikembangkan pada berbagai workshops (antara 1997-1999). Lebih dari 300 organisasi yang berkontribusi dalam proses modelling ini dan akhirnya

CRISP-DM 1.0 dipublikasikan pada 1999 (Irwan Budiman dkk., 2012).

Menurut Larose, pada bagian metode data mining ada 6 fase CRISP-DM (Cross Industry Standard Process for Data Mining) (Nugroho, 2014) yaitu:

a. Fase Pemahaman Bisnis (Business Understanding Phase)

Fase pemahaman bisnis bisa disebut sebagai fase pertama dengan kinerja yaitu memahami tujuan dan kebutuhan dari sudut pandang bisnis, kemudian menerjemahkan pengetahuan ini ke dalam pendefinisian masalah dalam data mining. Selanjutnya akan ditentukan rencana dan strategi untuk mencapai tujuan tersebut.

b. Fase Pemahaman Data (Data Understanding Phase)

Fase pemahaman data dimulai dengan pengumpulan data yang kemudian dilanjutkan dengan proses untuk mendapatkan pemahaman yang mendalam tentang data. Mengidentifikasi masalah kualitas data, atau untuk mendeteksi adanya bagian yang menarik dari data yang dapat digunakan untuk hipotesis untuk informasi yang tersembunyi.

c. Fase Pengolahan Data (Data Preparation Phase)

pada fase pengolahan data ini meliputi semua kegiatan untuk membangun dataset akhir (data yang akan diproses pada tahap pemodelan/modeling) dari data mentah. Tahap ini dapat diulang beberapa kali. Pada tahap ini juga mencakup pemilihan tabel, record, dan atribut-atribut data, termasuk proses pembersihan dan transformasi data untuk kemudian di jadikan masukan dalam tahap pemodelan.

d. Fase Pemodelan (Modeling Phase)

Pada fase pemodelan akan dilakukan pemilihan dan penerapan berbagai teknik pemodelan dan beberapa parameternya akan disesuaikan untuk mendapatkan nilai yang optimal. Secara khusus, ada beberapa teknik berbeda yang dapat

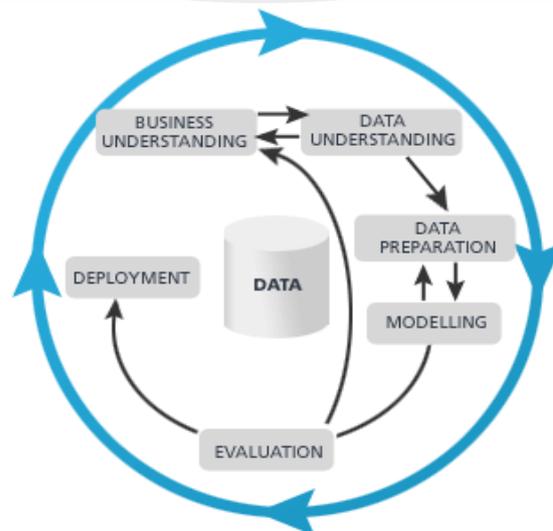
diterapkan untuk masalah data mining yang sama. Di pihak lain ada teknik pemodelan yang membutuhkan format data khusus. Sehingga pada tahap ini masih memungkinkan kembali ke tahap sebelumnya.

e. Fase Evaluasi (Evaluation Phase)

Pada fase evaluasi ini model sudah terbentuk dan diharapkan memiliki kualitas baik jika dilihat dari sudut pandang analisis data. Pada tahap ini akan dilakukan evaluasi terhadap keefektifan dan kualitas model sebelum digunakan dan menentukan apakah model dapat mencapai tujuan yang ditentukan pada fase awal (business understanding).

f. Fase Penyebaran (Deployment Phase)

Pada fase ini pengetahuan atau informasi yang telah diperoleh akan diatur dan dipresentasikan dalam bentuk khusus sehingga dapat digunakan oleh pengguna. Tahap deployment dapat berupa sebuah pembuatan laporan dengan sederhana atau mengimplementasikan proses data mining yang berulang dalam perusahaan. Dalam banyak kasus, tahap deployment melibatkan konsumen, disamping analisa data, karena sangat penting bagi konsumen untuk memahami tindakan apa yang harus dilakukan untuk menggunakan model yang telah dibuat.



**Gambar 2.1 Fase CRISP-DM (Larose, 2016)**

### 2.2.5 Algoritma K-Means

Dilihat dari latar belakang kata algoritma memiliki sejarah yang tidak pasti. Pada zaman dahulu orang-orang hanya menemukan kata Algorism. Algorism memiliki pengertian yaitu proses menghitung. ketika kita menggunakan hitungan maka kita akan disebut Algorism, maka kemudian ahli sejarah dalam bidang matematika mencari asal kata algoritma, sehingga ditemukannya asal mula kata dari nama penulis buku arab yang cukup terkenal Abu Ja'far Muhammad Ibnu Musa Al-Khawarizmi. Nama Al-Khawarizmi jika di baca orang barat berubah menjadi Algorism. sehingga kemudian lambat laun ketika Algorism dipakai sebagai suatu metode hitungan secara universal maka berubahlah makna katanya aslinya, jika di bawah di dalam bahasa Indonesia Algorism berubah menjadi Algoritma.

Pengertian Algoritma secara umum adalah suatu langkah-langkah strategis penyelesaian, masalah yang sistematis dan logis. Kata kunci algoritma adalah logis, karena semua kata yang di inputkan di dalam perhitungan harus logis dan harus di ketahui apakah bernilai salah atau benar, algoritma dikenal sebagai jantung dari komputer namun secara proses computer hanyalah pemroses yang bentuknya jika ditulis dalam notasi bahasa pemrograman maka akan disebut program.

Banyak para ahli yang berbeda pendapat tentang definisi K-means salah satunya Aryan. "menurut Aryan K-means merupakan bentuk Algoritma secara umum yang sering digunakan untuk meng clustering dokumen. Menurutnya prinsip awal dari K-means adalah menyusun K prototype atau pusat massa dari kumpulan data yang berdimensi N". menurut Chen yu (2010), pengertian K-means merupakan "algoritma yang digunakan untuk mengcluster objek berdasarkan atribut menjadi k partisi, di mana  $k < n$ ."

Dari pengertian yang dijabarkan oleh para ahli di atas, bisa ditarik kesimpulan untuk pengertian dari K-means adalah salah

satu metode pada data clustering non hirarki untuk mengcluster dokumen yang di mana agar kemudian bisa mempartisi data manual yang ada ke dalam bentuk satu atau lebih cluster berdasarkan atributnya untuk menjadi kepartisi, yang di mana  $k < n$ . K-means memiliki pengertian suatu Algoritma jika digunakan dalam pengelompokan secara partisi yang hasilnya akan membagi kelompok yang berbeda-beda dengan bentuk pemisahan data.

Algoritma K-means adalah algoritma yang banyak membutuhkan parameter-parameter masukan sebanyak  $k$  dan akan membagi kelompok  $n$  objek kedalam  $k$  cluster sehingga dapat mengetahui tingkat kemiripan anggota antara tingkat kemiripan anggota yang tinggi dan tingkat kemiripan anggota yang rendah. Untuk mengetahui kemiripan anggota maka kita harus mengukur kedekatan objek pada nilai mean atau disebut centroid cluster(Widyawati dkk., 2020).

Algoritma k-means merupakan suatu metode dan clustering non hirarki yang berusaha mempartisi data ke dalam bentuk satu atau lebih cluster/kelompok. Metode ini mempartisi ke dalam cluster/kelompok sehingga data yang memiliki karakteristik yang sama (high intra class similarity) dikelompokkan kedalam satu kluster yang sama dan yang memiliki karakteristik yang berbeda (low inter class similarity) dikelompokkan pada kelompok yang lain.

Proses clustering dimulaidengan mengidentifikasi data yang akan dikluster,  $X_{ij}$  ( $i=1, \dots, n$ ;  $j=1, \dots, m$ ) dengan  $n$  adalah jumlah data yang akan di kluster dan  $m$  adalah jumlah variabel. Pada awal iterasi, pusat setiap klaster ditetapkan secara bebas (sembarang),  $C_{kj}$  ( $k=1, \dots, k$ ;  $j=1, \dots, m$ ). kemudiandihitung jarak antara setiap data dengan setiap pusat cluster. Untuk melakukan penghitungan jarak data ke- $i$  ( $x_i$ ) pada pusat kluster ke- $k$  ( $c_k$ ), diberi nama ( $d_{ik}$ ), dapat digunakan formula Euclidean seperti pada persamaan berikut :

$$d_{ik} = \sqrt{\sum_{j=i}^m (X_{ik} - C_{ik})^2} \dots \dots \dots (1)$$

Suatu data akan menjadi anggota dari kluster ke-k apabila jarak data tersebut ke pusat kluster ke-k bernilai paling kecil jika dibandingkan dengan jarak ke pusat cluster lainnya. Hal ini dapat dihitung dengan menggunakan persamaan (2) di bawah ini, kemudian selanjutnya adalah kelompokan data-data yang menjadi anggota pada setiap kluster.

$$\min \sum_{k=1}^k d_{ik} = \sqrt{\sum_{j=1}^m (X_{ik} - C_{ik})^2} \dots \dots \dots (2)$$

Nilai pusat kluster yang baru dapat dihitung dengan cara mencari nilai rata-rata dari data-data yang menjadi anggota pada kluster tersebut, dengan menggunakan rumus pada persamaan (3) berikut ini :

$$C_{kj} = \frac{\sum_{i=1}^p X_{ij}}{P} \dots \dots \dots (3)$$

Di mana  $X_{ij}$ .kluster ke -k, P = banyak anggota kluster ke k. algoritma data dalam k-means adalah :

1. Tentukan jumlah kluster (k), tetapkan kluster sembarang.
2. Hitung jarak setiap data ke pusat cluster menggunakan persamaan (1).
3. Kelompokan data ke dalam kluster yang dengan jarak yang paling pendek menggunakan persamaan (2).
4. Hitung pusat cluster yang baru menggunakan persamaan (3).
5. Ulangi langkah dua sampai dengan empat hingga sudah tidak ada lagi data yang berpindah ke kluster yang lain.

## 2.2.6 R Studio

### 2.2.6.1 Pengertian R Studio

Menurut (Gio & Effendie, 2018) menyatakan R bukanlah program yang dapat Anda buka dan mulai gunakan, seperti Microsoft Word atau Internet Explorer. Sebaliknya, R adalah bahasa komputer, seperti C, C ++, atau UNIX. Di masa lalu, orang menjalankan kode R di jendela terminal UNIX seolah-olah mereka adalah peretas di film dari tahun 1980-an. Sekarang hampir semua orang menggunakan R dengan aplikasi yang disebut RStudio,

Menurut Andrie de Vries dan Joris Meys (Gio & Effendie, 2017) dalam bukunya yang berjudul R for Dummies. Menyatakan R lebih merupakan bahasa pemrograman daripada aplikasi. Saat Anda mengunduh R, Anda otomatis mengunduh aplikasi konsol yang sesuai untuk sistem operasi Anda. Tim Inti Pengembangan R telah berupaya keras agar R tersedia untuk berbagai jenis perangkat keras dan perangkat lunak. Ini berarti R tersedia untuk Windows, sistem Unix (seperti Linux), dan Mac. R itu sendiri adalah bahasa yang kuat yang melakukan berbagai fungsi, seperti manipulasi data, pemodelan statistik, dan grafik.”

Menurut Norman Matloff (Gio & Effendie, 2017) dalam bukunya yang berjudul The Art of R Programming. R merupakan bahasa scripting untuk manipulasi dan analisis data statistik. Menurutnya Mengapa menggunakan R Keutamaannya adalah :

- Implementasi domain publik dari bahasa statistik S yang dianggap luas; R / S adalah standar de facto di kalangan ahli statistik profesional.

- Dapat dibandingkan, dan seringkali lebih unggul, dalam kekuatan untuk produk komersial dalam arti yang paling.
- Tersedia untuk Windows, Mac, Linux.
- Selain mengaktifkan operasi statistik, ini adalah bahasa pemrograman umum, sehingga Anda dapat mengotomatiskan analisis Anda dan membuat fungsi baru.
- Struktur pemrograman berorientasi objek dan fungsional.
- Kumpulan data Anda disimpan di antara sesi, jadi Anda tidak perlu memuat ulang setiap kali.
- Sifat perangkat lunak terbuka berarti mendapatkan bantuan dari komunitas pengguna dengan mudah, dan banyak fungsi baru yang dikontribusikan oleh pengguna, banyak di antaranya adalah ahli statistik terkemuka. Saya harus memperingatkan Anda bahwa seseorang mengirimkan perintah ke R melalui teks, daripada klik mouse dalam Antarmuka Pengguna Grafis (GUI). Jika Anda tidak dapat hidup tanpa GUI, Anda harus mempertimbangkan untuk menggunakan salah satu GUI gratis yang telah dikembangkan untuk R, mis. Komandan R atau JGR. Perhatikan bahwa R pasti memiliki grafik banyak sekali. Tetapi grafik untuk keluaran, mis. plot, bukan untuk input.

R diciptakan (was created) oleh Ross Ihaka dan Robert Gentleman (nama depan sama-sama diawali dengan huruf R & R, Ross dan Robert) dari departemen statistika, di Universitas Auckland, New Zealand. Saat ini R dikembangkan oleh R Development Core Team. Munculnya R terinspirasi oleh bahasa statistika dengan nama S (Gio & Effendie,2017).

Dari hasil penjelasan dari beberapa tokoh diatas terkait R, dapat ditarik kesimpulan bahwa R merupakan bahasa pemrograman statistika yang sifatnya open-source dari pemrograman S dan bisa digunakan untuk analisis dan pemodelan statistika serta grafik.

RStudio merupakan salah satu bentuk frontend R yang cukup populer dan nyaman digunakan. Selain nyaman digunakan, RStudio memungkinkan kita melakukan penulisan laporan menggunakan Markdown atau Notebook serta membuat berbagai bentuk project seperti shiny, dll. Pada Rstudio juga memungkinkan kita mengatur working directory tanpa perlu mengetikkan sintaks pada Command, yang diperlukan hanya memilihnya di menu RStudio.



**Gambar 2.2 logo Rstudio**

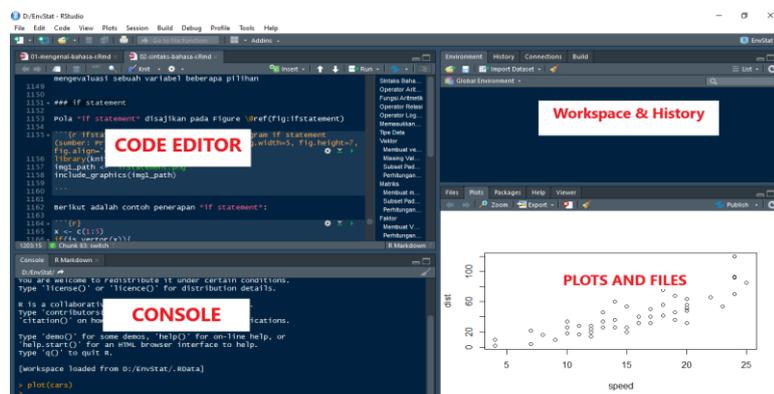
Menurut fiskal (2016) dalam wahyu dkk, ada beberapa keuntungan yang diberikan oleh R Studio, (Budiaji, 2019) yaitu :

1. Bantuan *code completion* yang akan menampilkan daftar perintah saat kita telah mengetik karakter pada command prompt.
2. Integrasi bantuan dalam satu layar, sehingga informasi hasil perintah *help* dapat langsung ditampilkan.

3. Tab *environment* yang menampilkan daftar objek yang di buat.
4. Tab *display files, plots, package* yang juga terintegrasi pada satu layar.
5. Menggunakan *shortcut* untuk posisi *cursor* masih dapat digunakan pada *command prompt*.

### 2.2.6.2 Tampilan Dan Perintah Menu Rstudio

Setelah R terinstal pada komputer maka dapat diakses melalui shortcut atau menu start. Prompt default “>” mengindikasikan bahwa R menunggu perintah yang diberikan, baik berupa fungsi maupun objek lain. R merupakan suatu bahasa berorientasi objek, artinya bahwa variabel, data, fungsi, hasil dan sebagainya, disimpan dalam memori aktif komputer dalam bentuk objek dan mempunyai sebuah nama. Pengguna dapat mengenakan aksi terhadap objek-objek tersebut melalui suatu operator (aritmatika, logika, dan perbandingan) dan fungsi (functions). Fungsi itu sendiri juga merupakan suatu objek. Penamaan terhadap objek bersifat case sensitive (membedakan huruf besar dan huruf kecil). Jadi X dan x merupakan objek yang berbeda (Nengsih dkk., 2020).



Gambar 2.3 Jendela Rstudio

R memiliki karakteristik yang berbeda dengan bahasa pemrograman lain seperti C++,python, dll. R memiliki aturan/sintaks yang berbeda dengan bahasa pemrograman yang lain yang membuatnya memiliki ciri khas tersendiri dibanding bahasa pemrograman yang lain, Beberapa ciri dan fitur pada R bersifat case sensitif. maksudnya adalah dalam proses input R huruf besar dan kecil sangat diperhatikan. Sebagai contoh kita ingin melihat apakah objek A dan B pada sintaks berikut:

```
A <- "Andi"  
B <- "andi"  
# cek kedua objek A dan B  
A == B  
## [1] FALSE  
# Kesimpulan : Kedua objek berbeda
```

### 2.2.7 Contoh Perhitungan K-means

Langkah- Langkah perhitungannya pada K-means di bawah ini:

1. Menentukan Jumlah cluster data
2. Tentukan titik pusat cluster
3. menghitung jarak objek
4. kelompokanobjek
5. jika kelompok data hasil perhitungan baru sama dengan hasil perhitungan kelompok data baru maka selesailah perhitungannya.

Diketahui : jumlah cluster = 2  
Jumlah data = 5  
Jumlah atribut = 2

Tabel 2.1 Contoh Data siswa

NO	Nama Siswa	Nilai B. Indonesia	Nilai B. Inggris
1	Mohammad wafiq izrol	79	75
2	Fitri indrianti	84	76
3	Deslina putrid	77	84
4	Candar maulana putra	78	86
5	Rita	82	82

**Iterasi pertama :**

- Menentukan pusat cluster.

Diambil data ke-2 sebagai pusat cluster ke-1                    84     75

Diambil data ke-5 sebagai pusat cluster ke-2                    82     82

- Menghitung jarak pusat cluster :

Untuk mengukur jarak antara dengan pusat *Cluster* digunakan *Euclidean Distance*, kemudian akan didapatkan matriks jarak yaitu C1 dan C2 sebagai berikut:

Rumus :

$$d(x,y) = |x - y| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

Tabel 2.2 hasil perhitungan

NO	Nama Siswa	C1	C2
1	Mohammad wafiq izrol	10.44	21.73
2	Fitri indrianti	0	24.02
3	Deslina putrid	24.15	14.56
4	Candar maulana putra	14.18	26.04
5	Rita	24.02	0

- Pengelompokan data : Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat cluster, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat cluster terdekat.

Tabel 2.3 hasil pengolahan data.

NO	Nama Siswa	C1	C2
1	Mohammad wafiq izrol	1	0
2	Fitri indrianti	1	0
3	Deslina putrid	0	1
4	Candar maulana putra	1	0
5	Rita	0	1

Hasil iterasi 1 :

Tabel 2.4 Hasil Iterasi pertama.

NO	Nama Siswa	Hasil
1	Mohammad wafiq izrol	Berada pada cluster 1
2	Fitri indrianti	Berada pada cluster 1
3	Deslina putrid	Berada pada cluster 2
4	Candar maulana putra	Berada pada cluster 1
5	Rita	Berada pada cluster 2

**Iterasi ke dua :**

Cluster 1 = 78.5; 76.2

Cluster 2 = 79.5; 83

Tabel 2.5 Iterasi Kedua

NO	Nama Siswa	C1	C2
1	Mohammad wafiq izrol	5.5	20.5
2	Fitri indrianti	7.9	22.9
3	Deslina putrid	21.1	7.2
4	Candar maulana putra	14.08	26.04
5	Rita	23.13	7.2

Langkah selanjutnya sama dengan perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antara data dengan pusat *cluster*, jarak ini menunjukkan bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat.

Tabel 2.6 Hasil Iterasi

NO	Nama Siswa	C1	C2
1	Mohammad wafiq izrol	1	0
2	Fitri indrianti	1	0
3	Deslina putrid	0	1
4	Candar maulana putra	1	0

Sehingga dengan terjadi persamaan maka tidak perlu lagi dilakukan perulangan, karena hasil cluster telah mencapai stabil dan konvergen. Pada penerapan k means di atas hanya sebuah contoh atau gambaran besar perhitungan k means yang dihitung untuk menemukan hasil cluster.

### 2.2.8 Flowchart (Diagram Alur)

Flowchart atau yang disebut diagram alur memiliki pengertian yakni penggambaran secara grafik dari langkah-langkah dan urutan-urutan prosedur dari suatu program. teknik analisis yang digunakan untuk menjelaskan aspek-aspek informasi secara jelas, tepat, dan logis. Bagan atau diagram alir menggunakan serangkaian simbol standar untuk menguraikan prosedur pengolahan transaksi yang digunakan oleh sebuah perusahaan, sekaligus menguraikan aliran data dalam sebuah system Ridlo, I. A. (2017).

Pada flowchart bisa diartikan sebagai urutan grafik yang memiliki langkah-langkah atau prosedur untuk menyelesaikan sebuah permasalahan sehingga bisa dikatakan bahwa flowsheet sangat membantu analisis dan program dalam memecahkan masalah dalam pengoperasian. Flowchart atau diagram alir memiliki simbol dan fungsi masing-masing, antara lain :

### 1. Flow Direction Symbol

Pada flow direction symbol memiliki fungsi dan mampu menghubungkan antara simbol pada flowchart sehingga flow direction symbol di sebut juga sebagai connection line.

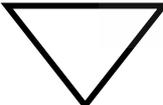
**Tabel 2.7 Simbol Penghubung Atau Alur**

Simbol	Nama Dan Fungsi
	<b>Simbol arus /flow :</b> menyatakan jalur arus proses.
	<b>Simbol communication link :</b> menyatakan transmisi data dari suatu lokasi ke lokasi lain.
	<b>Simbol connector :</b> menyatakan sambungan dari proses ke proses lainnya dalam halaman yang sama.
	<b>Simbol offlineconnector :</b> menyatakan sambungan dari proses ke proses lainnya dalam halaman yang berbeda

## 2. Processing symbol

Pada symbol processing ini menggambarkan operasi dalam suatu prosedur atau proses yang ada pada flowchart.

**Tabel 2.8 simbol Proses**

Simbol	Nama dan fungsi
	<b>Simbol proses</b> :menyatakan suatu tindakan (proses) yang ada pada flowchart.
	<b>Simbol manual</b> :untuk menggambarkan suatu proses yang di lakukan secara manual.
	<b>Simbol decision</b> :menunjukkan dua kondisi tertentu, untuk menentukan pilihan ya / tidak.
	<b>Simbol predefined process</b> :menyatakan penyediaan tempat penyimpanan suatu pengolahan untuk memberi harga awal.
	<b>Simbol terminal</b> : menyatakan permulaan atau akhir suatu program
	<b>Off-line storage</b> : menyatakan permulaan dan akhir suatu program atau proses.
	<b>Manual input</b> : menggambarkan proses input yang dilakukan secara manual
	<b>Input-output</b> :menggambarkan proses input ataupun output tanpa melihat jenis alat yang digunakan.
	<b>Punched card</b> : menggambarkan input yang berasal dari kartu atau output yang ditulis ke dalam media kartu

	<p><b>Magnetic-tape unit</b> : menggambarkan input yang bersumber dari pita magnetic atau output disimpan ke dalam pita magnetic</p>
	<p><b>Disk Storage</b> :menggambarkan input bersumber dari disk atau output yang di simpan di disk.</p>
	<p><b>Document</b> :menggambarkan proses pencetakan laporan ke printer.</p>
	<p><b>Display</b> :menggambarkan peralatan output yang di pakai merupakan layar.</p>

